Nijmegen School of Management
Department of Economics and Business Economics
Master's Thesis Economics (MAN-MTHEC)

Assessing the impact of the NFRD on sustainability in the EU

By Sander van Gruijthuijsen (S1044942)

Nijmegen, 30 June 2024

Program: Master's Program in Economics
Specialisation: Financial Economics
Supervisor: dr. S.C. Füllbrunn

Radboud Universiteit

Abstract

This study examines to what extent the Non-financial Reporting Directive (NFRD) improves the sustainable performance, i.e., the ESG scores, of companies within the European Union. It adds to the academic discipline by providing insights into the limited research that is conducted about the effects of the NFRD on the sustainable performance of companies within the EU. The study repeats and extends a previous study by mitigating its limitations and employing a more extensive methodology. A quasi-experimental research analysis through a difference-in-differences regression model based on matching was conducted. The results show insignificant differences between the treatment group and control group following the implementation of the NFRD and finds no support for a treatment effect for companies subject to the NFRD following its implementation. In conclusion, this study provides no support for a statistical improvement of sustainable performance, i.e., ESG scores, for companies within the EU following the NFRD. This research is restricted due to limited data on the pre-treatment period and ESG scores and financials of small companies. Recommendations for future research include examining the short- and long-term effects and the precursor impact of the NFRD and repeating this study for the newly enacted CSRD in the EU.

Keywords: ESG score(s), sustainability, Non-financial Reporting Directive (NFRD), difference-in-differences (DiD), matching

Table of Contents

1 Introduction

ESG has become a hot topic over the last decade and is part of a broader trend which is the growing importance of sustainability. These topics are particularly interesting for the financial domain as more and more companies are reporting on non-financial information. This results from companies voluntarily reporting on non-financial information or jurisdictions enacting laws of mandatory reporting on non-financial information. The main problem with voluntary reporting is the lack of standardization of metrics and the potential risk of greenwashing. The main problem with mandatory reporting is the collection, management, and verification of data, the complexity of the policy and the assessment of its effectiveness on sustainable performance.

Existing literature shows that mandatory disclosures are effective at improving the sustainability practices of companies in multiple ways. This improvement in sustainable performance following the implementation of the different directives is the case for each of the individual pillars of the ESG score as well as the overall ESG score (Chen et al., 2018; Jia & Chen, 2019; Lombardi et al., 2022; Ren et al., 2023). The main problem with the existing literature is that limited research has been conducted on the effects of the implementation of the Non-Financial Reporting Directive on the sustainable performance of companies within the European Union. Aluchna et al. (2022) did examine whether the Non-Financial Reporting Directive positively impacted the ESG performance and its pillar performance. However, their research faces some major limitations. These limitations include the limited dataset consisting of a small sample of exclusively Polish companies with an underdeveloped level of sustainability reporting and the small timeframe of the research.

This study examines the following research question: ‘to what extent does the Non-financial Reporting Directive improve the sustainable performance, i.e. the ESG scores, of companies within the European Union?’ The Non-financial Reporting Directive is named NFRD hereafter. This study addresses the problems mentioned above by extending the dataset from a small sample of Polish companies with an underdeveloped level of sustainability reporting to a larger sample of companies across the European Union which includes companies with an ex-ante developed level of sustainability reporting. Furthermore, the timeframe is extended from a six-year period of which three years belong to the pre- and post-treatment period to a twelve-year period of which six years belong to the pre- and post-treatment period. Lastly, the methodology is improved by using a matching technique that matches the treatment units to the control units on a set of covariates that is in line with previous literature (Chen et al., 2018; Ren et al., 2023; Yuan & Wu, 2024).

This study employs a difference-in-differences model based on matching which controls for a multitude of control variables and covariates. This model belongs to the quasi-experimental design studies. Difference-in-differences is a statistical technique that estimates the effect of the interaction term of two binary variables, one indicating whether the period belongs to the pre- or post-treatment period and the other indicating whether a company is a treatment or control unit. Matching is a technique that mitigates potential endogeneity by matching treatment units to control units based on their similarity of characteristics to achieve balance in these characteristics between the two groups. The control variables and covariates that are controlled for are in line with previous research (Yuan & Wu, 2024). The matching technique utilized in this research is 1:1 cardinality matching.

The pre-treatment period amounts to the years 2011-2016 and the post-treatment period amounts to the years 2017-2022. Treatment units are companies subject to the NFRD, i.e., large public-interest entities operating in the European Union with more than 500 employees. Control units are companies operating in the European Union which are not subject to the directive, i.e., companies which are not considered to be large public-interest entities and have 500 employees or less. Allocating companies based on their number of employees to different groups could cause problems as a potential treatment effect could be the result of size differences between the treatment group and control group. The study mitigates this potential problem by controlling for confounding factors that could explain the treatment difference. Controlling for confounding factors is done by matching the treatment units and control units on a set of covariates. The panel dataset consists of 7,205 observations from 1,190 different companies. 700 observations belong to control units and 6,505 observations belong to treatment units. After matching, the dataset consists of 700 treatment units and 700 control units.

This research finds no support for a treatment effect for the companies within the EU following the implementation of the NFRD, and thereby finds no support for a statistical improvement of sustainable performance, i.e., ESG scores, for companies within the EU following the NFRD. This contrasts with the hypotheses described hereafter which predicted a positive effect for all pillar scores and the overall ESG score and thus provides contradicting results to the existing literature.

The paper is structured as follows. Chapter 2, Literature Review, formulates hypotheses based on a theoretical background of academic knowledge to make empirical predictions. Methodology is the third chapter of the paper and is divided into multiple subsections. These include Data and sample, Variables, and Modeling. These subsections serve the purpose of elaborating on the design of the research by describing the choices and assumptions made. Results is the fourth chapter of the research and is also divided into multiple subsections. These subsections entail Descriptive statistics, Hypothesis testing, and Robustness check. This chapter summarizes the characteristics of the dataset, provides empirical findings to test the hypotheses and examines whether these findings are robust. Chapter 5, Conclusion and discussion, finalizes the research by offering an analysis and interpretation of the empirical findings. Furthermore, it discusses both the theoretical and practical implications and the potential limitations of the research.

2 Literature Review

To examine whether the NFRD significantly improved the ESG performance of companies within the European Union, a number of hypotheses are formulated. These hypotheses make empirical predictions based on insights embedded in academic literature. The literature mentioned hereafter first provides a theoretical foundation of the influence of reporting standards on the social responsibility and behaviour of a company. After the theoretical foundation, empirical findings from the literature examining the effect of various reporting standards on the sustainability of companies are assessed. The literature review is finalized with the formulation of the hypotheses.

Chen et al. (2018) state that mandatory CSR disclosure impacts the activities of a firm as the increased transparency following the mandatory disclosure eases the ability of governments and interest groups to put pressure on firms for more CSR activities. Furthermore, other stakeholders also closely watch and act upon negative CSR incidents. The reasoning is that these stakeholders benefit from the increased transparency following the CSR disclosure and as a result can identify firms that they can pressure to increase CSR activities. Chen et al. (2018) also state that mandatory CSR disclosure may facilitate monitoring of the environmental performance by external parties and thereby incentivizing firms to decrease pollution levels or, if firms do not focus on pollution levels, improve social externalities by focusing on staff protection and public relations. They conclude that political and social stakeholder pressure related to disclosure regulation can cause a firm to take action, and thereby has at least some benefit to society. They also conclude that spending on CSR is mainly driven by political and social factors instead of economic factors (Chen et al., 2018).

Ren et al. (2023) summarize from previous literature that both the quantity and quality of CSR disclosure significantly improved following the implementation of mandatory disclosure regulations, showing a high-level conformity. Mandatory disclosure regulation helps managers to understand the beliefs and expectations of their stakeholders, thereby influencing the beliefs of the firm on the expected appropriateness and to adhere to the current behaviour of the expectations of stakeholders. Furthermore, mandatory disclosure regulation increases transparency significantly, thereby reducing information asymmetries between firms and external stakeholders and increasing the possibility of rewards or punishments by others for the environmental performance of the firm (Ren et al., 2023).

Jackson et al. (2020) support the findings by Chen et al. (2018) and Ren et al. (2023) as they find that firms operating in countries that require mandatory non-financial disclosure take on significantly more CSR activities, while having the largest impact on firms with previously low levels of CSR activities. Christensen et al. (2021) go beyond the insights mentioned above and argue that firms reduce or even disinvest certain activities which are negatively viewed by stakeholders or are related to highly sensitive CSR matters. They also state that a change in firm behaviour following corporate disclosure is more likely to follow from mandatory reporting than from voluntary disclosures (Christensen et al., 2021).

Research shows that after the implementation of mandatory CSR disclosure by China in 2008 cities that are most impacted by the mandate experience a decrease in industrial wastewater and lower sulfur dioxide (SO2) emission levels as firms pollute less (Chen et al., 2018). Furthermore, the same CSR disclosure mandate also resulted in substantially higher green innovation for mandatory CSR reporting firms compared to non-CSR reporting firms (Ren et al., 2023). Another paper found a decrease in the city-day air pollution following the establishment of the central environmental protection inspection (CEPI) in China in 2016 (Jia & Chen, 2019). As a decrease in the amount of waste and emissions and higher green innovation are all components of the Environmental pillar and impact that pillar positively, this results in an improved Environmental score and thus an overall improved ESG score.

Chen et al. (2018) also show that firms subject to the Chinese 2008 mandatory CSR disclosure experienced fewer workplace fatalities after the disclosure. This shows that the health & safety of the workforce is enhanced. As Workforce is a factor of the Social pillar, fewer workplace fatalities improve the Social pillar, which in turn improves the overall ESG score.

The Governance pillar is the final pillar of ESG. A study shows that following the NFRD by the EU large public companies exercised improved transparency and better organization of sustainable, environmental and societal information (Lombardi et al., 2022). This is supported by the argument made that mandatory CSR disclosure increases the transparency of the company substantially, which in turn reduces information asymmetries between both firms and external stakeholders (Ren et al., 2023). These effects improve the CSR strategy as a factor of the Governance pillar through an improvement of ESG reporting and transparency. An enhancement of Governance in turn results in an improved ESG score.

Aluchna et al. (2022) empirically found an improvement in overall ESG score, environmental performance, and social performance for Polish companies subject to the NFRD following the implementation of the directive. This study extends on the research of Aluchna et al. (2022) by overcoming their limitations. The limitations amount to the limited dataset consisting of a small sample of exclusively Polish companies with an underdeveloped level of sustainability reporting, the small timeframe of the research, and not controlling for confounding effects. This research mitigates these limitations by extending the dataset from a small sample of Polish companies with an underdeveloped level of sustainability reporting to a larger sample of companies across the European Union which includes companies with an ex-ante developed level of sustainability reporting. Furthermore, the timeframe is extended from a six-year period of which three years belong to the pre- and post-treatment period to a twelve-year period of which six years belong to the pre- and post-treatment period. Lastly, this research controls for differences in size between the treatment units and control units as confounding factors.

To examine the effect of the NFRD on the ESG performance of companies within the EU and thereby giving answer to the research question specified previously, a set of concise hypotheses is formulated. The hypotheses are formulated this way to achieve clear and testable hypotheses. The hypotheses described hereafter are derived from the critical reflection of existing academic literature above consisting of a theoretical foundation and empirical findings.

Hypothesis 1: The NFRD has a positive effect on the ESG performance of companies within the European Union

Hypothesis 1.1: The NFRD has a positive effect on the environmental performance of companies within the European Union

Hypothesis 1.2: The NFRD has a positive effect on the social performance of companies within the European Union

Hypothesis 1.3: The NFRD has a positive effect on the governance performance of companies within the European Union

3 Methodology

3.1 Data and sample

To examine whether the NFRD by the European Union positively impacted the sustainable activities, i.e., the ESG scores, of companies operating within the EU this research compares two groups based on being subject to the NFRD. The two groups are big companies, companies with more than 500 employees, which are subject to the directive and small companies, companies with 500 or fewer employees, which are not subject to the directive. The differences between the two groups are examined with a difference-in-differences regression model. The regression model consists of two dummy variables, one indicating whether the company is subject to the NFRD and the other indicating whether the period is post-treatment, the interaction term of the two dummy variables, and a number of covariates. The model controls for confounding effects by using a matching technique that matches the two groups on a set of covariates as explained in Chen et al. (2018) and Ren et al. (2023). The covariates which are matched on follow from previous literature (Yuan & Wu, 2024). Data on the ESG scores as the dependent variables are derived from LSEG Datastream, while data on the financial variables, company board variables, and GDP variable as the covariates are derived from the LSEG Datastream, BoardEx, and OECD databases, respectively. This study repeats the research of Aluchna et al. (2022) by examining the effect of the NFRD on the sustainability, i.e. ESG scores, of companies within the EU and extends on their research by overcoming their limitations. This is done by extending the set of companies, extending the timeframe, and controlling for a greater number of confounding factors.

To examine the effect of the NFRD on the sustainable performance of companies operating within the EU a portion of companies is taken of the total amount of companies subject to the directive. The companies which are subject to the NFRD are large public-interest entities operating in the European Union with more than 500 employees (Hahnkamper-Vandenbulcke & European Parliamentary Research Service, 2021). Therefore, the treatment group in this research also consists of companies which adhere to the same criteria. The control group consists of companies operating in the European Union which are not subject to the directive, i.e., companies which are not considered to be large public-interest entities and have 500 employees or less. The country of domicile as an ISO code is utilized to determine whether a company is operating within the EU and to determine the respective country, and thus the GDP levels, the company belongs to. The United Kingdom left the European Union halfway through the post-treatment period on 1 February 2020 (European Commission, n.d.-b). Therefore, the United Kingdom is no longer an EU member state from 2020 onwards and the observations of companies operating in the United Kingdom are excluded from this analysis from 2020 onwards.

The financial year 2022 is the latest finished financial year for which financial information is available to its full extent as there are several companies for which the financial year 2023 has not ended yet and still must report on their financials for that year. Companies within the European Union subject to the NFRD had to report in line with the directive for the first time in 2018, for the 2017 financial year (Hahnkamper-Vandenbulcke & European Parliamentary Research Service, 2021). This results in a six-year period from 2017 up to and including 2022 for which the directive was operating and for which financial information is available. To examine whether the directive had a significant impact on the sustainability of companies within the European Union, this study examines the differences in ESG scores of the treatment group and control group in the six-year period before and after the directive coming into effect. As a result, the financial years 2011-2016 form the pre-treatment period and the financial years 2017-2022 form the post-treatment period to be examined.

The observations of a particular company for a single year are excluded from the analysis when the data for a variable is missing. Subsequently, the total dataset of this research consists of 7,205 observations from 1,190 different companies. Out of the 7,205 observations, 2,185 observations amount to the pre-treatment period, i.e., the years 2011 up to and including 2016 and 5,020 amount to the post-treatment period, i.e., the years 2017 up to and including 2022. Furthermore, out of the more than 7,000 observations, 700 belong to control units and 6,505 belong to treatment units.

The set of companies is acquired from LSEG Datastream (formerly known as Refinitiv Workspace/Eikon). The datasets utilized are ‘Market Europe’ and ‘FTSE All-Share Index’. Data on the variables is collected from multiple databases. The databases include LSEG Datastream, BoardEx, and OECD. Of these databases LSEG Datastream is predominantly used to retrieve data.

LSEG Datastream is used to collect data on the ESG and financial variables. BoardEx is utilized to gather data on company board variables and is accessible through WRDS. The OECD database is used to retrieve data on the levels of GDP of the respective country the company is operating in.

The dataset consists of panel data as there is a multitude of companies and a sequence of years subject to the research. The two statistical computing and graphics programming languages used to conduct research are Stata and R. Stata is used to create the dummy variables, create the financial and company board variables out of a raw dataset and combine several datasets to create a panel dataset. R is used to create the regression model and conduct the analysis.

3.2 Variables

The dependent variables in this research include the ESG score (*ESG*) and its subcomponents the Environmental (*ENV*), Social (*SOC*) and Governance (*GOV*) scores. These scores are obtained from LSEG. The ESG scores from LSEG are based on 186 metrics grouped in 10 different categories of three separate pillars. LSEG assigns each metric, category and pillar a certain weight depending on the relevance to the respective industry of the company to ultimately compute an ESG score between 0 and 100. The Environmental pillar consists of the categories: Emission, Innovation, and Resource use; the Social pillar consists of the categories: Community, Human rights, Product responsibility, and Workforce; the Governance pillar consists of the categories: CSR strategy, Management, and Shareholders (LSEG, 2023).

The independent variables in this research are the variables *Post* and *Treatment*. The *Post* variable is a dummy variable that equals 1 when the year belongs to the post-treatment period of the directive. This is the case for the years 2017 up to and including 2022. The dummy variable receives a value equal to 0 for the years 2011-2016, i.e., the pre-treatment period. The *Treatment* variable is a dummy variable that receives value 1 when a company within the European Union is subject to the NFRD, i.e., the company has more than 500 employees. If a company within the European Union has 500 or fewer employees, the dummy variable amounts to 0.

The control variables and covariates that are controlled for in this study are composed of various firm-level characteristics and one national indicator from the benchmark paper by Yuan and Wu (2024). The firm-level characteristics include the company size (*Size*), liabilities to assets ratio (*Leverage*), return on assets (*ROA*), CEO duality (*Duality*), profit or loss (*Profit*), ratio of independent board members (*Independence*), shareholding ratio of top shareholder (*Top1*), presence of Big Four accounting firm (*Big4*), public ownership (*SOE*), operating income growth (*Growth*), Tobin's Q (*TobinsQ*), and company age (*Age*) (Yuan & Wu, 2024). The national indicator is the GDP level (*lnGDP*). Yuan and Wu (2024) use another indicator to account for the region the company operates in by dividing the regional secondary industry out value by the regional GDP. This variable is omitted in this study as the respective industries the companies operate in are not of interest to this research. To account for the different currencies across the European Union all financial data is converted to US dollars. Table 1 provides a concise overview of the abovementioned variables and the descriptions of these variables.

TABLE 1. DESCRIPTION OF THE VARIABLES

VARIABLE	DESCRIPTION
ESG	ESG score from LSEG Datastream
ESGC	ESG Combined score from LSEG Datastream
ENV	Environmental (E) score from LSEG Datastream
SOC	Social (S) score from LSEG Datastream
GOV	Governance (G) score from LSEG Datastream
POST	Dummy variable equal to 1 for the financial years [2017, 2022], 0 otherwise
TREATMENT	Dummy variable equal to 1 for EU companies with > 500 employees, 0 otherwise
SIZE	Ln (Total Assets)
LEVERAGE	Total Liabilities/Total Assets
ROA	ROA variable from LSEG Datastream
DUALITY	Dummy variable equal to 1 if the CEO is also the chairman, 0 otherwise
PROFIT	Dummy variable equal to 1 if the net profit > 0, 0 otherwise
INDEPENDENCE	Ratio of independent board members to the total number of board members
TOP1	Shareholding ratio of the biggest shareholder at the end of the year
BIG4	Dummy variable equal to 1 if company is audited by a Big Four accounting firm*, 0 otherwise
SOE	Dummy variable equal to 1 if company is state-owned, 0 otherwise
GROWTH	(Operating income of the current year/operating income of the previous year) - 1
TOBIN'S Q	Tobin's Q from LSEG Datastream
AGE	Ln (years since IPO)
GDP	Ln (national GDP)

Notes: * Deloitte, EY (Ernst & Young), KPMG and PwC (PricewaterhouseCoopers) are considered to be the four biggest accounting firms.

3.3 Modeling

This study performs an empirical research analysis by examining the effects of the enactment of the NFRD on the sustainability, i.e., the ESG scores, of companies within the EU. To conduct the empirical research analysis a difference-in-differences model based on matching is developed. This research design excludes randomization across groups and therefore, belongs to the quasi-experimental design studies.

Difference-in-differences is a statistical technique that estimates the effect of a treatment by comparing the change in results over a certain time period of the treatment group to the change in results over a certain time period of the control group. As mentioned previously, the treatment group in this research consists of large public-interest entities within the EU, i.e., companies and groups with more than 500 employees. The control group consists of non-large public-interest entities within the EU, i.e., companies and groups with 500 employees or less. Difference-in-differences is accounted for by creating two dummy variables, one indicating whether a company is a treatment unit (value = 1) or a control unit (value = 0) and the other indicating whether the firm-year is from 2017-2022 and thus post-treatment (value = 1) or whether the firm-year is 2011-2016 and thus pre-treatment (value = 0).

Matching is used to estimate the causal effect of a binary treatment variable while controlling for measured pre-treatment variables, typically confounding variables (Greifer, 2023c). It matches treatment units to control units based on a set of covariates as first introduced by Rosenbaum and Rubin (1983). It is a method that is used to mitigate potential endogeneity by taking into account the control variables and covariates that could explain the treatment difference (Ren et al., 2023). The matching method that is used to match the treatment units to the control units is called 1:1 cardinality matching. A balance tolerance level of 0.1 is assigned to the model. This matching technique is a pure subset selection method (Greifer, 2023b).

Before settling on the 1:1 cardinality matching as the way to match the control units to the treatment units multiple matching techniques were tried. These include but are not limited to Nearest Neighbor Matching, Optimal Full Matching, Optimal Pair Matching, Exact Matching, and Subclassification. Cardinality yielded the best balance. The improved balance after matching is shown in detail in Appendix A and Appendix C. After matching, all standardized mean differences for the covariates were within the tolerance level of 0.1 indicating adequate balance. 1:1 cardinality matching matches treatment units to control units on a one-to-one basis. Considering the imbalance of the 700 control units and 6,505 treatment units this leaves units unmatched as clearly shown in Appendix B. No units were discarded because of the matching.

To test for the hypotheses described above the effect of the treatment is estimated by fitting a linear regression model which includes the matching weights. G-computation is used to estimate the average treatment effect in the matched sample (ATM). To account for heteroscedasticity, the delta method estimates the variance of the model by using cluster-robust standard errors with pair membership as the clustering variable (Greifer, 2023b). The difference-in-differences model based on matching is analyzed by the following regression:

$$(1)\quad Y = \beta_0 + \beta_1 Post + \beta_2 Treatment + \beta_3 Post \times Treatment + \beta_4 Controls + \varepsilon$$

The regression above shows that the dependent variable *Y* is regressed on the dummy variables *Post* and *Treatment*, the interaction term *Post x Treatment*, and *Controls*. The dependent variable *Y* takes the form of multiple scores each regressed independently. These scores include the ESG score (*ESG*) and its subcomponent scores: *ENV*, *SOC*, *GOV*. The dummy variables and the control variables are described in detail in section 3.2 Variables. The variable of interest in the regression above is the coefficient β_3 of the interaction term (*Post * Treatment*). The error term ε indicates the unexplained variance in ESG scores from the difference-in-differences model. The robustness check that are tested for are further specified in section 4.3 Robustness check.

4 Results

4.1 Descriptive statistics

Table 2 below provides the descriptive statistics of the total dataset split by the binary Treatment variable. The first noticeable difference between the treatment and control units is the amount of observations N for every variable. The table shows that the dataset consists of 6,505 treatment units and 700 control units. As 90.28% of the total units are treated units and 9.72% are control units, this indicates an imbalance in the dataset.

The variables *ESG*, *ESGC*, *ENV*, *SOC*, and *GOV* all have a theoretical distribution of values between 0 up to and including 100. The table shows that for both the treatment units and control units there are units with values close to the upper bound and the lower bound of these variables. This could indicate a large spread of values. This is supported by the considerable standard deviation of all these variables for both the treatment and control units with a higher standard deviation indicating a larger deviation from the mean. Furthermore, the table does not indicate that the values of the *ESGC* variable exceed the values of the *ESG* variable for a firm-year observation. This is in line with the methodology about the ESG Combined score as explained in section 4.3 Robustness check.

The values of the *Post* variable indicate that for the treatment units 68.9% (0.689) of the observations amount to the post-treatment period and 31.1% amount to the pre-treatment period. For the control units 76.7% (0.767) of the observations amount to the post-treatment period, while 23,3% amounts to the pre-treatment period.

The values of the *ROA* variable in the table show that this variable could deviate to quite some extent with considerable values for the minimum and maximum of both the treatment and control units and the standard deviation of the control units.

The mean for the *Duality* variable is noticeably higher for the treatment units compared to the control units. This variable is a dummy variable equal to 1 if the CEO is also the chairman of the company. This indicates that the companies belonging to the treatment units on average experience more duality compared to the control units, i.e., there is a higher chance for the CEO to also be the chairman of a company when the company is subjective to the directive.

The higher average of the *Big4* dummy variable for the treated units compared to the control units indicates that on average treatment units are more likely to be audited by a Big Four accounting firm.

The last noticeable differences are within the *Growth* and *TobinsQ* variables. The minimum and maximum values for *Growth* show that, taking the mean and standard deviation for these variables into account, there are significant outliers for both the treatment and control units. Outliers are also present in the maximum values of the *TobinsQ* for both the treatment and control units. Furthermore, the standard deviation of the same *TobinsQ* variable is noticeable higher for the control units than for the treatment units.

TABLE 2. DESCRIPTIVE STATISTICS OF THE VARIABLES BY TREATMENT VARIABLE PRIOR TO MATCHING

Descriptive statistics by Treatment variable										
	Treatment units					Control units				
Variable	N	Mean	St. Dev.	Min	Max	N	Mean	St. Dev.	Min	Max
ESG	6,505	59.562	17.528	1.750	95.720	700	43.305	20.392	3.910	91.140
ESGC	6,505	57.128	16.626	1.750	93.860	700	43.244	20.379	3.910	91.140
ENV	6,505	57.264	24.592	0.000	98.910	700	38.112	27.986	0.000	96.150
SOC	6,505	64.040	20.266	1.330	98.470	700	47.042	21.539	3.300	95.680
GOV	6,505	55.038	21.335	1.530	98.560	700	42.383	23.564	2.970	97.140
Post	6,505	0.689	0.463	0	1	700	0.767	0.423	0	1
Size	6,505	15.766	1.778	11.179	21.885	700	14.509	1.274	10.010	17.941
Leverage	6,505	0.620	0.198	0.004	1.615	700	0.332	0.236	-0.302	0.988
ROA	6,505	5.693	7.514	-53.220	221.720	700	8.185	25.118	-84.600	269.110
Duality	6,505	0.305	0.461	0	1	700	0.084	0.278	0	1
Profit	6,505	0.889	0.314	0	1	700	0.811	0.391	0	1
Independence	6,505	0.775	0.143	0.143	1.000	700	0.813	0.154	0.333	1.000
Top1	6,505	0.269	0.209	0.002	0.998	700	0.224	0.182	0.013	0.983
Big4	6,505	0.913	0.282	0	1	700	0.844	0.363	0	1
SOE	6,505	0.030	0.171	0	1	700	0.004	0.065	0	1
Growth	6,505	-0.127	10.034	-504.403	212.744	700	-0.433	10.817	-218.874	68.566
TobinsQ	6,505	1.823	1.588	0.478	58.866	700	2.163	6.215	0.392	91.201
Age	6,505	2.829	0.929	0.000	5.106	700	2.613	0.932	0.000	4.357
lnGDP	6,505	7.237	0.990	3.883	8.361	700	7.326	0.900	4.247	8.361

Table 3 below provides the descriptive statistics of the matched dataset split by the binary Treatment variable. The first noticeable is the amount of observations N for the Treatment units. The amount of Treatment units decreased from 6,505 in the total dataset, i.e., prior to matching, to 700 units in the matched dataset. This is the result of 1:1 cardinality matching as the matching technique for this analysis as previously stated in section 3.3 Modeling. The sample sizes are shown in Table 11 in Appendix B.

Comparing the averages for the *ESG*, *ESGC*, *ENV*, *SOC*, and *GOV* variables for the treatment units prior to matching and after matching shows that the averages for all these variables have decreased after matching. This indicates that after matching the dependent variables for the treatment units are, on average, more similar to the dependent variables for the control units as the value of the means are closer in value to each other.

The averages of the covariates (*Size, Leverage, ROA, Duality, Profit, Independence, Top1, Big4, SOE, Growth, TobinsQ, Age, lnGDP*) of the treatment units provide further support that the matching procedure achieved balance on the covariates between the treatment units and control units. This follows from the observation that the averages of the covariates for the treatment units are noticeably more in line with the averages of the covariates for the control units after matching than is the case prior to matching. This observation applies to every single covariate.

TABLE 3. DESCRIPTIVE STATISTICS OF THE VARIABLES BY TREATMENT VARIABLE FOR THE MATCHED DATA

Descriptive statistics matched dataset by Treatment variable										
	Treatment units					Control units				
Variable	N	Mean	St. Dev.	Min	Max	N	Mean	St. Dev.	Min	Max
ESG	700	52.885	18.339	2.540	93.170	700	43.305	20.392	3.910	91.140
ESGC	700	51.516	17.585	2.540	93.170	700	43.244	20.379	3.910	91.140
ENV	700	47.660	25.502	0.000	97.940	700	38.112	27.986	0.000	96.150
SOC	700	56.755	21.205	1.790	98.470	700	47.042	21.539	3.300	95.680
GOV	700	51.918	21.499	2.540	94.410	700	42.383	23.564	2.970	97.140
Post	700	0.694	0.461	0	1	700	0.767	0.423	0	1
Size	700	14.686	1.458	11.179	18.954	700	14.509	1.274	10.010	17.941
Leverage	700	0.352	0.145	0.004	1.210	700	0.332	0.236	-0.302	0.988
ROA	700	8.904	11.064	-27.460	80.130	700	8.185	25.118	-84.600	269.110
Duality	700	0.121	0.327	0	1	700	0.084	0.278	0	1
Profit	700	0.784	0.412	0	1	700	0.811	0.391	0	1
Independence	700	0.799	0.146	0.250	1.000	700	0.813	0.154	0.333	1.000
Top1	700	0.245	0.206	0.002	0.863	700	0.224	0.182	0.013	0.983
Big4	700	0.864	0.343	0	1	700	0.844	0.363	0	1
SOE	700	0.020	0.140	0	1	700	0.004	0.065	0	1
Growth	700	-0.532	13.758	-347.462	31.569	700	-0.433	10.817	-218.874	68.566
TobinsQ	700	2.034	1.577	0.478	12.813	700	2.163	6.215	0.392	91.201
Age	700	2.706	0.988	0.000	4.779	700	2.613	0.932	0.000	4.357
lnGDP	700	7.393	0.919	3.984	8.361	700	7.326	0.900	4.247	8.361

4.2 Hypothesis testing

Table 4 below provides the average treatment effects in the matched sample (ATM) for the different dependent variables (DV) tested for. These dependent variables include the ESG score (*ESG*) and its subcomponent scores: *ENV*, *SOC*, *GOV*. The table shows a slightly negative estimate of 0.242 for *ESG* and 0.240 for *GOV*, a negative estimate of 2.817 for *SOC*, while a positive estimate of 2.763 is shown for *ENV*. These results would indicate that companies subject to the directive experience a slightly smaller improvement in ESG score and governance performance, a smaller improvement in social performance, and a bigger improvement in environmental performance after the implementation of the NFRD compared to companies who are not subject to the NFRD over the same period.

However, the table also shows that for all dependent variables, insignificant probability values of 0.909, 0.332, 0.243, and 0.925 are found. As a result, insignificant differences between the treatment units and control units before and after the implementation of the directive are found for the dependent variables. Thus, no significant (treatment) effects are found following the implementation of the NFRD.

TABLE 4. ATM FOR THE DIFFERENT DEPENDENT VARIABLES

DV	Interaction term	Estimate	Std. Error	z	Pr(>|z|)	S	2.5%	97.5%
ESG	PostxTreatment 1 - 0	-0.242	2.117	-0.114	0.909	0.137	-4.391	3.908
ENV	PostxTreatment 1 - 0	2.763	2.846	0.971	0.332	1.593	-2.815	8.341
SOC	PostxTreatment 1 - 0	-2.817	2.411	-1.168	0.243	2.043	-7.542	1.908
GOV	PostxTreatment 1 - 0	-0.240	2.555	-0.094	0.925	0.112	-5.247	4.768

4.3 Robustness check

To establish the robustness of the findings of this study, an alternative measure of the ESG score is deployed in the regression model. This alternative measure is the ESG Combined score (ESGC) and replaces the regular ESG score as the dependent variable in this section. The ESG Combined score (ESGC) is a weighted average of the ESG score and the ESG controversies score. This combined score provides a more comprehensive evaluation of a company's sustainability impact and conduct over time by discounting the ESG performance of companies (LSEG, 2023). This is done by covering the impact of significant, material ESG controversies with recent controversies reflected in the latest completed fiscal period (LSEG, 2023). The ESG controversies score is a score based on 23 ESG controversy data points that penalize a particular company when it is involved in a scandal.

Furthermore, the controversies score also addresses the market cap bias which large companies are typically subject to. The market cap bias explains that large-cap companies are influenced disproportionately large by their market capitalization by the market and thereby also attract more media attention compared to smaller-cap companies. As a result, these large-cap companies also suffer disproportionately large from significant, material ESG controversies. The controversies score accounts for the disproportionality of the market cap bias by incorporating severity weights, which assign different weights to a company based on the market capitalization of that particular company (LSEG, 2023).

When the ESG controversies score is greater than or equal to the ESG score, i.e., the company does not face any significant negative media stories during the fiscal period, the ESG score equals the ESGC score. When the ESG controversies score is below the ESG score, and thus controversies are present for a company during a fiscal period, the ESGC score is the average of the ESG score and the ESG controversies score (LSEG, 2023). To examine the ESGC score as a robustness check, the following regression is developed:

$$\text{(2)} \quad ESGC = \beta_0 + \beta_1 Post + \beta_2 Treatment + \beta_3 Post \times Treatment + \beta_4 Controls + \varepsilon$$

Table 5 below shows the average treatment effect in the matched sample (ATM) with the ESGC score as the dependent variable. The results from the table indicate that companies subject to the directive experience a slightly bigger improvement in ESGC performance after the implementation of the NFRD compared to companies who are not subject to the NFRD over the same period. This follows from the positive estimate of 0.095. However, the table also shows that for the ESGC score an insignificant probability value of 0.964 is found. As a result, an insignificant difference between the treatment units and control units before and after implementation of the directive is found for the ESGC score.

The robustness check provides contradicting evidence to the results of testing the hypotheses in section 4.2 in the sense that a positive estimate is found for the ESG Combined score, while a negative estimate is found for the ordinary ESG score. However, the estimates are considered to be relatively low and relatively close in value to each other and the probability value shows an insignificance for the treatment effect on the ESGC score.

Therefore, the robustness check does not indicate to deviate from the interpretations and conclusions drawn from the main analysis of testing the hypothesis previously as for both insignificant (treatment) effects are found.

TABLE 5. ATM OF ESGC SCORE

DV	Interaction term	Estimate	Std. Error	z	Pr(>|z|)	S	2.5%	97.5%
ESGC	PostxTreatment 1 - 0	0.095	2.121	0.045	0.964	0.052	-4.063	4.252

5 Conclusion and discussion

This research examines the effect of the NFRD on the sustainable performance of companies operating within the EU. The effect is examined by performing a difference-in-differences analysis on the ESG scores of the matched samples of control units and treatment units. The results show negative estimates for the *ESG*, *SOC*, and *GOV* scores and positive estimates for the ENV and ESGC scores. These results would indicate that a positive treatment effect is present for the treatment companies after the implementation of the NFRD for the environmental and ESGC performance. However, the results also indicate that these potential treatment effects in the *ENV* and *ESGC* scores are insignificant. Insignificant effects are also found for the negative estimates of the *ESG*, *SOC*, and *GOV* scores. Although treatment effects are found for some aspects of the ESG score, these treatment effects are not consistently found for every dependent variable and no statistically significant difference is observed over time between the two groups for any of the dependent variables. This research finds no support for a treatment effect for the companies within the EU following the implementation of the NFRD, and thereby finds no support for the hypotheses posed previously. In conclusion, this study provides no support for a statistical improvement of sustainable performance, i.e., ESG scores, for companies within the EU following the NFRD.

These findings contrast with the hypotheses described earlier. The hypotheses embedded in the literature review formulated positive effects for all scores, while a positive effect is exclusively found for the environmental score and ESGC score. Furthermore, these positive effects are of insignificant influence which is also in contrast to the literature. A possible reason for the inconsistency between the formulated hypotheses and the results in this study could be related to the differences within the matched samples. Treatment units and control units differ from each other due to their size, which makes them incomparable. This study mitigates this problem by matching the units on their similarity of a set of covariates. However, it is likely that the matching is not perfect as differences continue to persist and crucial confounding factors are potentially left out of the matching progress. Therefore, it could be that the difference in ESG scores between the treatment units and the control units is explained by the difference in covariates in contrast to the treatment effect. This is a potential explanation for the statistically insignificant differences.

Yet, this explanation does not justify the contradicting results and conclusions of the study by Aluchna et al. (2022) as they did not account for confounding factors in their study while finding significant positive treatment effects following the implementation of the NFRD. A potential explanation for the contradiction is that due to the omission of confounding factors by Aluchna et al. (2022) their significantly positive treatment effects are the result of the differences in size between the treatment companies and control companies. Section 4.1 Descriptive statistics shows that treatment companies, typically larger companies, experience, on average, higher ESG scores compared to control companies, typically smaller companies. The differences in averages of ESG scores are greater before matching than after matching. Furthermore, larger companies are more prone to the market cap bias as explained in section 4.3 Robustness check and could therefore be more induced to improve their operations or reporting in such a way to behave more sustainably and as a result score higher in terms of ESG performance. Combining these considerations could explain that the differences in size between treatment units and control units are potentially the confounding factors for the treatment effects found by Aluchna et al. (2022), which they do not control for. This study matches the treatment units to the control units on a set of confounding factors, thereby controlling for size differences between the units and as a result finds no significant effects following the implementation of the NFRD.

A limitation of this research relates to the dataset. The availability of data for ESG scores for years prior to the implementation of the NFRD is limited. This limited availability is also the case for ESG scores and financials of relatively small companies not subject to the directive in general. As a result, the generalizability of the effect of the directive to all companies within the European Union is also limited as the treatment effect of the directive could deviate with a larger sample. This limitation has been mitigated as much as possible by utilizing the largest datasets available within LSEG Datastream that provide data on companies operating within the European Union and the United Kingdom. Nevertheless, it cannot be excluded that the generalizability of the treatment to the whole of the European Union is limited to some extent due to the limited data availability.

Another constraint of this research concerns the methodological choices of the study. It is up for debate whether the control and treatments are comparable as differences between the two units are present because of their size. This poses the question whether a treatment effect could be derived from the incomplete comparability. The methodological approach employed controls for these confounding factors and fully mitigates the incomparability by matching the control and treatment units on a set of firm-level and regional covariates that is in line with previous literature to mitigate potential endogeneity. Nonetheless, it cannot be ruled out that the units are not truly comparable as potentially crucial confounding variables are omitted from the analysis, and thus, differences in size between the two units will continue to persist.

Future studies could extend this research by examining the short- and long-term effects of the implementation of the NFRD. This research focused on whether the enforcement of the NFRD impacted the sustainable performance of companies within the European Union by comparing the ESG scores of treatment companies with the scores of control companies between the pre- and post-treatment period. The effect of the implementation of the NFRD could be made more comprehensive by examining whether the effect deviates for different timeframes during the treatment period. Yuan and Wu (2024) point out that even before the implementation of the Environmental Protection Tax Law by China in 2018, the enactment of the law in 2016 would likely have had a precursor impact on the sustainable performance of the companies. Furthermore, Aluchna et al. (2022) suggest that it requires time for companies to introduce transparency principles into their organizational structures. This follows from their conclusion that there is a delay in sustainable performance improvement as they found improvements in the sustainable performances of Polish companies a few years after the implementation of the NFRD and not for the year following the implementation of the directive. Future research could provide more in-depth knowledge of the effect by examining whether the precursor impact and the differences in short- and long-term effects are also present in the sustainable performance of companies across the EU following the implementation of the NFRD.

Another way for future studies to extend on this research is by repeating this study for the newly enacted sustainability directive, the CSRD (European Commission, n.d.-a). By examining the effects of the CSRD in the same light as the NFRD in this study and drawing on their differences and similarities in outcome a more comprehensive understanding of the effects of sustainability reporting is achieved.

6 Bibliography

Aluchna, M., Roszkowska-Menkes, M., & Kaminski, B. (2022). From talk to action: the effects of the non-financial reporting directive on ESG performance. *Meditari Accountancy Research*, *31*(7), 1-25. https://doi.org/10.1108/medar-12-2021-1530

Chen, Y. C., Hung, M. Y., & Wang, Y. X. (2018). The effect of mandatory CSR disclosure on firm profitability and social externalities: Evidence from China. *Journal of Accounting & Economics*, *65*(1), 169-190. https://doi.org/10.1016/j.jacceco.2017.11.009

Christensen, H. B., Hail, L., & Leuz, C. (2021). Mandatory CSR and sustainability reporting: economic analysis and literature review. *Review of Accounting Studies*, *26*(3), 1176-1248. https://doi.org/10.1007/s11142-021-09609-5

European Commission. (n.d.-a). Corporate sustainability reporting. Finance. https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en#legislation

European Commission. (n.d.-b). The EU-UK withdrawal Agreement. https://commission.europa.eu/strategy-and-policy/relations-non-eu-countries/relations-united-kingdom/eu-uk-withdrawal-agreement_en

Greifer, N. (2023a, October 12). Assessing balance. https://cran.r-project.org/web/packages/MatchIt/vignettes/assessing-balance.html

Greifer, N. (2023b, October 12). Estimating effects after matching. https://cran.r-project.org/web/packages/MatchIt/vignettes/estimating-effects.html

Greifer, N. (2023c, October 12). Matching methods. https://cran.r-project.org/web/packages/MatchIt/vignettes/matching-methods.html\

Greifer, N. (2023d, October 12). MatchIt: Getting started. https://cran.r-project.org/web/packages/MatchIt/vignettes/MatchIt.html

Hahnkamper-Vandenbulcke, N. & European Parliamentary Research Service. (2021). Non-financial reporting directive. In EPRS (Report PE 654.213). https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/654213/EPRS_BRI(2021)654213_EN.pdf

Jackson, G., Bartosch, J., Avetisyan, E., Kinderman, D., & Knudsen, J. S. (2020). Mandatory Non-financial Disclosure and Its Influence on CSR: An International Comparison. *Journal of Business Ethics*, *162*(2), 323-342. https://doi.org/10.1007/s10551-019-04200-0

Jia, K., & Chen, S. W. (2019). Could campaign-style enforcement improve environmental performance? Evidence from China's central environmental protection inspection. *Journal of Environmental Management*, *245*, 282-290. https://doi.org/10.1016/j.jenvman.2019.05.114

Lombardi, R., Cosentino, A., Sura, A., & Galeotti, M. (2022). The impact of the EU Directive on non-financial information: Novel features of the Italian case. *Meditari Accountancy Research*, *30*(6), 1419-1448. https://doi.org/10.1108/medar-06-2019-0507

LSEG. (2023). Environmental, social and governance scores from LSEG. https://www.lseg.com/content/dam/data-analytics/en_us/documents/methodology/lseg-esg-scores-methodology.pdf

Ren, S. G., Huang, M., Liu, D. H., & Yan, J. (2023). Understanding the Impact of Mandatory CSR Disclosure on Green Innovation: Evidence from Chinese Listed Firms. *British Journal of Management*, *34*(2), 576-594. https://doi.org/10.1111/1467-8551.12609

Rosenbaum, P. R., & Rubin, D. B. (1983). THE CENTRAL ROLE OF THE PROPENSITY SCORE IN OBSERVATIONAL STUDIES FOR CAUSAL EFFECTS. *Biometrika*, *70*(1), 41-55. https://doi.org/10.1093/biomet/70.1.41

Yuan, K., & Wu, B. Z. (2024). The crowding-out effect of the environmental regulation on corporate sustainability. *Frontiers in Environmental Science*, *11*, 16, Article 1273278. https://doi.org/10.3389/fenvs.2023.1273278

7 Appendix A. Summary of balance before and after matching

Table 6 and Table 7 below show the summary of the balance for the unmatched and matched data, respectively. The "Means Treated'' and ''Means Control'' columns display the averages of the covariates for the treatment group and the control group, respectively. The "Std. Mean Diff.'' column indicates the standardized mean differences between the treatment group and control group. "Var. Ratio" shows the variance ratios. The abbreviation ''eCDF'' in the last two columns represents the empirical cumulative distribution function. The last two columns show the averages and the maximum values of this statistic for every variable.

Absolute values of the standardized mean difference ''Std. Mean Diff.'' close to 0 indicate good balance, with 0.1 and 0.05 recommended as values for prognostically important covariates (Greifer, 2023a). A variance ratio "Var. Ratio" close to 1 indicates good balance, with values between 0.5 and 2 as the recommended values. Values of the ''eCDF'' statistics range from 0 to 1, with values closer to 0 indicating better balance (Greifer, 2023a).

Comparing the difference in statistics between the dataset prior to matching and matched dataset shows that matching improved the balance noticeably. This follows from the observation that for the standardized mean differences and eCDF statistics the values after matching are closer to their recommended values compared to the dataset prior to matching.

TABLE 6. SUMMARY OF BALANCE FOR DATA PRIOR TO MATCHING

	MEANS TREATED	MEANS CONTROL	STD. MEAN DIFF.	VAR. RATIO	ECDF MEAN	ECDF MAX
SIZE	15.766	14.509	0.707	1.946	0.204	0.313
LEVERAGE	0.620	0.332	1.457	0.705	0.318	0.465
ROA	5.693	8.185	-0.332	0.090	0.075	0.130
DUALITY	0.305	0.084	0.480		0.221	0.221
PROFIT	0.889	0.811	0.248		0.078	0.078
INDEPENDENCE	0.775	0.813	-0.269	0.858	0.061	0.177
TOP1	0.269	0.224	0.214	1.312	0.049	0.135
BIG4	0.913	0.844	0.244		0.069	0.069
SOE	0.030	0.004	0.152		0.026	0.026
GROWTH	-0.127	-0.433	0.030	0.861	0.050	0.097
TOBINSQ	1.823	2.163	-0.214	0.065	0.178	0.323
AGE	2.829	2.613	0.233	0.994	0.037	0.090
LNGDP	7.237	7.326	-0.090	1.210	0.035	0.108

Table 7 shows that good balance has been achieved for the matched dataset. This follows from the observation that the standardized mean differences for all covariates are within the tolerance threshold of 0.1, most of the values of the variance ratios are between 0.5 and 2, and a great amount of the values of the eCDF statistics are relatively close to 0. All these observations indicate that good balance has been achieved for the matched dataset.

TABLE 7. SUMMARY OF BALANCE FOR MATCHED DATA

	MEANS TREATED	MEANS CONTROL	STD. MEAN DIFF.	VAR. RATIO	ECDF MEAN	ECDF MAX
SIZE	14.686	14.509	0.100	1.309	0.036	0.081
LEVERAGE	0.352	0.332	0.100	0.376	0.074	0.241
ROA	8.904	8.185	0.096	0.194	0.044	0.107
DUALITY	0.121	0.084	0.081		0.037	0.037
PROFIT	0.784	0.811	-0.086		0.027	0.027
INDEPENDENCE	0.799	0.813	-0.100	0.898	0.029	0.101
TOP1	0.245	0.224	0.100	1.271	0.032	0.093
BIG4	0.864	0.844	0.071		0.020	0.020
SOE	0.020	0.004	0.092		0.016	0.016
GROWTH	-0.532	-0.433	-0.010	1.618	0.026	0.069
TOBINSQ	2.034	2.163	-0.081	0.064	0.221	0.390
AGE	2.706	2.613	0.100	1.124	0.019	0.074
LNGDP	7.393	7.326	0.067	1.043	0.020	0.101

8 Appendix B. Sample sizes before and after matching

Table 8 below shows the sample sizes of the treatment group and the control group before and after matching. It shows that the total dataset amounts to 700 control units and 6,505 treatment units, while the matched dataset amounts to 700 control units and 700 treatment units. The effective sample size (ESS) in the table below equals the number of units for the respective groups. The matching technique 1:1 cardinality matching left 5,805 treatment units unmatched as it matches treatment units to control units on a one-to-one basis. The matching discarded no units.

TABLE 8. SAMPLE SIZES BEFORE AND AFTER MATCHING

	CONTROL	**TREATED**
ALL (ESS)	700	6,505
ALL	700	6,505
MATCHED (ESS)	700	700
MATCHED	700	700
UNMATCHED	0	5,805
DISCARDED	0	0

9 Appendix C. Absolute standardized mean differences before and after matching

Figure 1 below shows the absolute standardized mean differences between the treatment group and control group for the total dataset ("All") and the matched dataset ("Matched"). The figure is a graphical visualization of the standardized mean difference between the treatment group and control group before matching, indicated by the white dots, and after matching, indicated by the black dots. The figure shows that the overall absolute standardized mean differences have decreased noticeably after matching and that the mean difference is lower for every covariate for the matched dataset compared to the total dataset. Furthermore, the graph shows that the absolute standardized mean difference for every covariate is within the 0.1 threshold indicating good balance. This follows logically from the matching technique 1:1 cardinality matching with the balance tolerance set at 0.1.

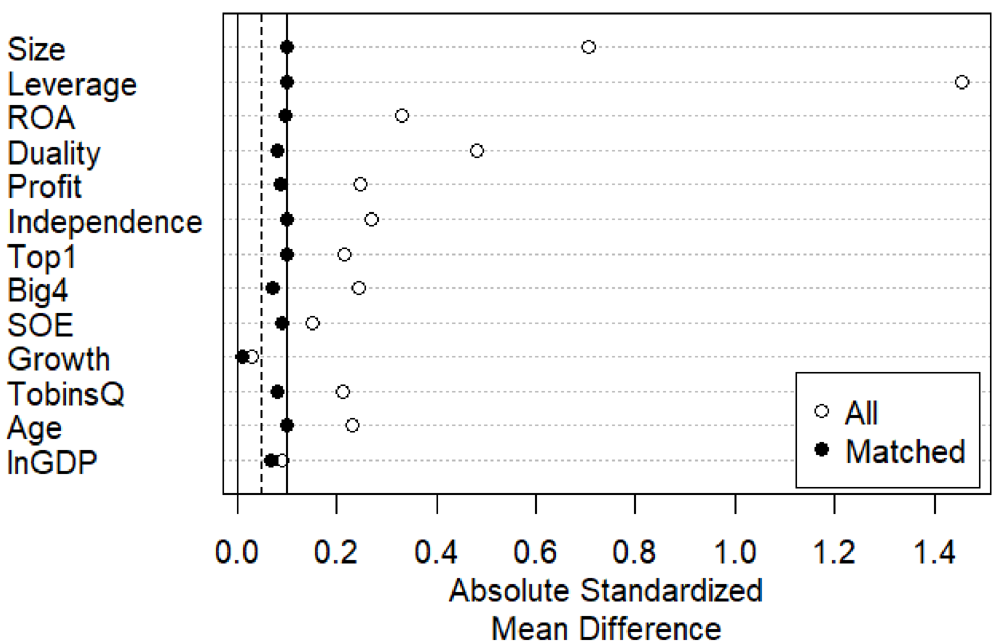

FIGURE 1. LOVE PLOT ON DISTRIBUTION OF ABSOLUTE STANDARDIZED MEAN DIFFERENCES OF COVARIATES FOR ALL AND MATCHED DATA